



**Manchester  
Metropolitan  
University**

---

Ainsworth, Stephanie, Welbourne, Stephen and Hesketh, Anne (2016) Lexical restructuring in preliterate children: Evidence from novel measures of phonological representation. *Applied Psycholinguistics*, 37 (04). pp. 997-1023. ISSN 0142-7164

---

**Downloaded from:** <https://e-space.mmu.ac.uk/621748/>

**Version:** Published Version

**Publisher:** Cambridge University Press

**DOI:** <https://doi.org/10.1017/S0142716415000338>

**Usage rights:** Creative Commons: Attribution 3.0

Please cite the published version

<https://e-space.mmu.ac.uk>

# Lexical restructuring in preliterate children: Evidence from novel measures of phonological representation

STEPHANIE AINSWORTH  
*Manchester Metropolitan University*

STEPHEN WELBOURNE and ANNE HESKETH  
*University of Manchester*

Received: March 4, 2014      Accepted for publication: June 6, 2015

## ADDRESS FOR CORRESPONDENCE

Stephanie Ainsworth, Faculty of Education, Brooks Building, Manchester Metropolitan University, Bonsall Street, Manchester M15 6GX, UK. E-mail: [s.ainsworth@mmu.ac.uk](mailto:s.ainsworth@mmu.ac.uk)

## ABSTRACT

There is substantial debate in the literature surrounding the development of children's phonological representations (PRs). Although infant studies have shown children's representations to contain fine phonetic detail, a consensus is yet to be reached about how and when phonemic categories emerge. This study used novel implicit PR measures with preschool children ( $n = 38$ , aged 3 years, 6 months to 4 years, 6 months) to test predictions made by these competing accounts of PR development. The measures were designed to probe PR segmentation at the phoneme (rather than the phone) level without requiring an explicit awareness of phonemes. The results provide evidence in support of vocabulary driven restructuring, with PR segmentation found to be related to vocabulary when controlling for age.

Phonological representations (PRs) store information about the sounds in words that a person holds in his or her vocabulary. Although PRs appear within many models of speech processing (e.g., Jusczyk, 1993; Metsala & Walley, 1998; Werker & Curtin, 2005), there is still much to learn about how their structure might change from infancy to adulthood (e.g., Walley, 1993; Walley, Metsala, & Garlock, 2003). It is important for us to understand the development of PRs and the processes driving their maturation because PRs play a key role in both early speech development and the later process of learning to read (Fowler, 1991; Metsala & Walley, 1998; Swan & Goswami, 1997).

Within the current paper the term PRs is used to describe the way that children store phonological word forms in memory. There is continuing debate within the literature about whether or not children store word forms separately for use in perception and production (e.g., Diehl, Lotto, & Holt, 2004; Hickok & Poeppel, 2007; Liberman & Mattingly, 1985; Mackensen & Fikkert, 2010). While speech perception and speech production have traditionally been studied as separate domains (Casserly & Pisoni, 2010), there are numerous sources of evidence that suggest that the two processes are closely related (for a review, see Casserly & Pisoni, 2010) and may draw upon common representational structures (Galantucci, Fowler, & Turvey, 2006; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985).

Evidence for close links between speech perception and production comes from a range of empirical sources (Casserly & Pisoni, 2010; Galantucci et al., 2006) including the phenomenon of phonetic convergence, where speakers adjust their speech during conversation to minimize the difference between their own productions and those of the other speaker (Pardo, 2006) and the “McGurk effect,” which describes how people’s perception of a heard sound can be overridden by the mouth movements of the speaker (Massaro, 1987; McGurk & MacDonald, 1976). Although alternative explanations have been proposed that do not require speech perception and speech production to share common representations (e.g., for alternative account of the McGurk effect, see Massaro, 1998), it has been argued that phonological representations that mediate both processes provide the most parsimonious explanation (Galantucci et al., 2006) for these effects.

In support of this view, neurobiological studies have discovered mirror neurons in monkeys (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) and humans (Rizzolatti, Fadiga, Matelli, et al., 1996) that become activated both when an action is performed and when an action is observed. This finding suggests an equivalence, or at least a direct connection, between perception and production at the representational level (Casserly & Pisoni, 2010). Further supporting evidence comes from a transcranial magnetic stimulation study that found that a person’s motor plans become activated when the person hears someone else speak (Fadiga, Craighero, Buccino & Rizzolatti, 2002). In light of the growing body of evidence in favor of unified representation across production and perception, we have assumed within the current work that children store the sound structure of words within a common set of phonological representations, recruited for both speech perception and production.

Another key term used within this paper is *phonological awareness* (PA), which is used to describe children’s ability to reflect upon or manipulate the sounds within words (e.g., can children tell you that the word “cat” can be segmented into the phonemes /k/-/æ/-/t/?). One of the key aims of this paper was to devise a battery of PR measures that allow us to probe the development of children’s underlying representations, independent of their explicit awareness of phonological structure (PA).

## THEORIES OF PR DEVELOPMENT

Theories of PR development have historically fallen into one of two camps: the accessibility view and the emergent view (for a review, see Walley et al., 2003).

The accessibility view proposes that children's PRs are segmented from infancy but that children are only able to access the phonemic segments explicitly once children learn to read and/or develop sufficient metacognitive skills (Liberman, Shankweiler, & Liberman, 1989; Morais, Bertelson, Cary, & Alegria, 1986). Evidence in support of the accessibility view comes from studies showing infants to be sensitive to minimal contrasts in perceptual tasks from as early as 14 months (e.g., Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Swingley, 2009; Swingley & Aslin, 2000, 2002) alongside studies with school children showing a sudden surge in performance on phoneme awareness tasks around the time that children learn to read (for a review, see Ziegler & Goswami, 2005). Taken together, these studies have been interpreted as evidence that infants have adultlike representations from the outset but that they only become conscious of their phonological structure through mastery of the alphabetic principle (Liberman et al., 1989; Morais et al., 1986).

In contrast, the emergent view proposes that infants' PRs are initially underspecified, becoming gradually restructured as children's vocabularies grow (Fowler, 1991; Metsala & Walley, 1998; Walley, 1993). The rationale behind this view is that infants do not need to store much detail in their representations to be able to discriminate between known words when their vocabularies are small, but as their vocabularies expand, children's PRs need to become better specified to allow efficient representation of similar-sounding words. The refinement of children's PRs is proposed to take place gradually, continuing (for some words) until as late as 7 years of age (Fowler, 1991; Metsala & Walley, 1998; Nittrouer, Studdert-Kennedy, & McGowan, 1989).

One theory under the umbrella of the emergent view, the lexical restructuring model (LRM), proposes that children's PRs are initially holistic, becoming gradually segmented at the onset-rime (e.g., cat → /k/-/æt/) and then the phoneme level (cat → /k/-/æ/-/t/) as children's vocabularies grow. The LRM predicts that high-frequency words from dense neighborhoods (those for which there are many similar-sounding words already in the lexicon) will be restructured first, with some less frequent words from sparse neighborhoods remaining underspecified into adulthood. Within the LRM the degree of restructuring of PRs constrains the child's explicit knowledge of the sound system, referred to as PA (Metsala & Walley, 1998). The LRM proposes that segmented PRs are prerequisite for the development of PA and predicts that PA will follow the developmental path of PR restructuring, with a delay between the two (Metsala & Walley, 1998).

Another variant of the emergent view, psycholinguistic grain size theory (Ziegler & Goswami, 2005), shares with the LRM the idea that vocabulary growth drives restructuring, but differs in terms of the extent to which PRs can be restructured in the absence of literacy. The LRM suggests that phonemic representations precede phonemic awareness (Metsala & Walley, 1998), whereas grain size theory proposes that PRs will only become phonemic once children have been taught about the correspondence between phonemes and graphemes (Ziegler & Goswami, 2005). Within this account, literacy precedes phonemic representation, with children's PRs becoming stored in terms of individual sounds only after they have been taught about phoneme-grapheme correspondences through the process of learning to read.

Over the last decade, there has been growing support for an alternative theoretical framework that encompasses elements of both the accessibility and the emergent accounts. Within processing rich information from multidimensional interactive representations (PRIMIR), rich, detailed information from the speech signal is stored from infancy but the degree to which infants can access different levels of this information changes according to the child's developmental level and the task at hand (Werker & Curtin, 2005). Phoneme-level representation is proposed to emerge gradually within the phoneme plane as statistical regularities are extracted from word-level exemplars. Fuzzy phonemelike categories start to form during infancy but do not become adultlike until they are "sharpened up" through literacy experience. The PRIMIR framework was born out of a need to reconcile seemingly contradictory studies where infants of the same age appear to have different levels of phonological specificity depending on which task was used. For example, while 14-month-old infants appear to have access to fine levels of phonetic detail on mispronunciation detection tasks (Swingley & Aslin, 2002), they struggle to distinguish between similar-sounding words during word learning tasks (Werker, Fennell, Corcoran, & Stager, 2002). In contrast, 20-month-old infants have been shown to use fine levels of phonetic detail successfully on both tasks (Werker et al., 2002). Within PRIMIR this differential performance by task is explained in terms of dynamic filters that modulate access to information across three interactive planes (the general perceptual, word form, and phoneme planes). The degree to which infants access different levels of the rich information stored from the input is influenced by three filters: the infants' initial biases (e.g., sensitivity to their native language), developmental level, and the demands of the task (Werker & Curtin, 2005).

## PHONOLOGICAL REPRESENTATION, VOCABULARY GROWTH, AND ROLE OF LITERACY

The theoretical accounts outlined above make different predictions with regard to two key questions:

1. Are infants' phonological representations qualitatively different from adults'?

According to the accessibility position, infants' representations are adultlike from the beginning (Liberman et al., 1989; Morais et al., 1986). Each of the other three accounts discussed above (the LRM, psycholinguistic grain size theory, and PRIMIR), however, suggests some kind of representational change over development. While the LRM proposes a gradual increase in segmentation of children's PRs from large to smaller grain sizes (Metsala & Walley, 1998), psycholinguistic grain size theory frames the restructuring process in terms of details being added at all grain sizes as children learn more words (Ziegler & Goswami, 2005). Within PRIMIR, although rich phonetic detail is present from the outset, the representation of phonemes is proposed to emerge gradually.

Numerous studies have set out to test these predictions empirically; however, a consensus has yet to be reached. Evidence in support of a developmental shift in the way phonological representations are stored comes from differential performance on a range of speech perception tasks (for a review, see Metsala & Walley, 1998).

For example, while adults tend to categorize pseudowords in terms of phonemic similarity (Treiman & Breaux, 1982), children tend to categorize them in terms of global similarity (how similar they sound overall). This finding has been interpreted as evidence of children's PRs being less segmented than adults', with a whole-to-parts shift proposed to take place gradually over childhood (Metsala & Walley, 1998). By contrast, several infants studies (as mentioned above) have shown infants' representations to contain sufficient detail to allow them to detect minimal contrasts, leading some researchers to conclude that infants' PRs are similar in nature to adults'. However, others have argued that early sensitivity to minimal contrasts does not require infants' representations to be stored in terms of phonemes. Ziegler and Goswami (2005) point out that there is often confusion within the literature with regard to whether infants are sensitive to phonemes or phones. In other words, just because an infant can discriminate between *ped*-*bed* this does not necessarily mean that they categorize the *p* in *pick* as the same phoneme as the *p* in *cup*. There is, therefore, a need for measures that probe children's representations at the phoneme rather than the phone level.

2. If phonological representations do change over development, to what extent is representational change driven by vocabulary growth?

Within the two key variants of the emergent account discussed above (the LRM, see Metsala & Walley, 1998; and psycholinguistic grain size theory, see Ziegler & Goswami, 2005), vocabulary growth is predicted to drive an increase in phonological specificity. Within the LRM, this restructuring is framed in terms of PRs becoming both increasingly segmented and increasingly detailed. We would, therefore, expect vocabulary growth to be related to both PR accuracy and PR segmentation. Literacy is not afforded a specific role within the LRM: vocabulary growth is emphasized as the key driver of lexical restructuring, although Metsala and Walley (1998) acknowledge that literacy experience may impact on PR development.

Psycholinguistic grain size theory frames restructuring slightly differently in terms of detail being added to PRs at all grain sizes rather than in terms of a whole-to-parts transition. Within psycholinguistic grain size theory, words will only become stored in terms of phonemes once children gain knowledge of phoneme-grapheme correspondences when learning to read. Grain size theory, therefore, predicts vocabulary size to be related to PR accuracy and literacy experience to be related to PR segmentation.

Finally, PRIMIR predicts that both vocabulary growth and literacy will be related to the emergence of phonemic representation (Werker & Curtin, 2005). As children learn more words, statistical regularities within the word form plane will become more apparent allowing phonemelike categories to form within the phoneme plane. Literacy is then proposed to complete the process, refining the emerging phonemes into an adultlike form. PRIMIR therefore predicts both vocabulary growth and literacy experience to be drivers of PR segmentation. However, it does not predict a direct relation between either vocabulary or literacy experience and PR accuracy, given that children's representations are presumed to contain detailed information from early infancy. Performance on PR accuracy measures would, however, be expected to vary depending on the developmental ability of the child and the demands of the particular task.

A number of infant studies have investigated the relationship between phonological representation and vocabulary size yielding mixed results. Although Werker et al. (2002) found significant correlations between vocabulary size and word learning within a sample of 14-month-old infants, no such correlation was found when the same task was used with older infants aged 20 months. Swingley and Aslin (2000) also found no correlation between vocabulary and sensitivity to mispronunciations in word recognition with infants aged 18 to 23 months. As mentioned above, these infant studies measure sensitivity to differences at the phone rather than the phoneme level. Within both the LRM and PRIMIR it is at the phoneme level (as well as at the onset-rime level in the case of the LRM) that we would expect to find a correlation between vocabulary growth and phonological representation. There is, therefore, a need for studies that look for the correlation between vocabulary growth and phonemic representation predicted by the LRM and PRIMIR.

To date, the majority of research has focused either on the very early stages of development (infants aged 14–24 months or on the PRs of children aged 4 and above; e.g., Claessen, Heath, Fletcher, Hogben, & Leitao, 2009; Foy & Mann, 2009). However, the intermediate preschool period that extends beyond infancy but precedes the point where children are exposed to formal literacy instruction is of particular theoretical interest. Within variants of the emergent account and PRIMIR, representational change is predicted to be a protracted process (Metsala & Walley, 1998; Werker & Curtin, 2005; Ziegler & Goswami, 2005), and so we would expect to find correlations between vocabulary size and measures of phonological specificity beyond infancy. In order to investigate the degree to which lexical restructuring takes place before children learn explicitly about phoneme–grapheme correspondences in school, and to thus separate the potential influences of vocabulary growth and literacy, we need to test children before they acquire this knowledge. While infant studies are limited in terms of the range of methodologies that are suitable in terms of metacognitive demands, preschool samples provide a potential “window” where children have not yet learned to read but are at an age where they can access more abstract tasks. By constructing tasks appropriate for preschool children that probe phonological representation at the phoneme level, we can look for evidence of emerging phonemic representation (rather than sensitivity to phones) before children begin formal literacy instruction. This is the aim of the current study.

## MEASUREMENT OF PHONOLOGICAL REPRESENTATIONS

The measurement of PRs is challenging for two reasons. First, many PR tasks require explicit phonological awareness that draws on metacognitive abilities rather than purely on the representations themselves. When trying to measure underlying PRs, we need to create tasks that do not require explicit PA, especially when we are measuring the PRs of preschool children whose metacognitive abilities are not fully developed. For this reason, tasks need to be as implicit as possible and should not require children to explicitly reflect upon or manipulate phonemes. Second, measurement of underlying representation necessarily involves behavioral tasks that inevitably involve other elements of speech processing and that may impede



interpretation of results. Measures of PRs that require a spoken response are potentially confounded by output difficulties (Claessen & Leitao, 2012; McNeill & Hesketh, 2010). In other words, if a child fails to produce a word correctly, we do not know if this is because the child has an inaccurate representation of the word, inaccurate motor programs, problems with articulation, or all three. In order to avoid the confound of speech output difficulties, it is necessary to use at least some tasks that do not require a spoken response, for example, receptive PR judgment tasks (Claessen & Leitao, 2012, McNeill & Hesketh, 2010).

Another way to circumvent this problem is to isolate a particular aspect of speech processing through comparisons across different tasks. For example Vance, Stackhouse, and Wells (2005) compared children's performance on naming, word repetition, and nonword repetition using the same stimuli across different age groups. Vance et al. found that while older children showed an advantage for repeating words over nonwords, 3-year-olds performed equally well for word and nonword repetition. Vance et al. (2005) suggest that this may be an indication that young children process both words and nonwords using a "bottom up" approach, relying increasingly on lexical information as they get older. The current study attempts to replicate this intriguing result.

Tasks designed to measure children's underlying PRs fall broadly into two categories: measures of PR accuracy/accessibility and measures of PR segmentation.

#### *Measures of PR accuracy and accessibility*

The first category aims to measure how accurately children store the sounds in words and how quickly they can retrieve them. The most commonly used tasks within this category are naming, word repetition, articulation correction (Anthony et al., 2010; Fowler & Swainson, 2004), in which the child is asked to correct a puppet's mispronunciation, and mispronunciation detection. The mode of delivery for mispronunciation detection varies according to age: whereas older children are asked directly whether a stimulus is pronounced correctly (e.g., Anthony et al., 2010; Claessen et al., 2009), mispronunciation detection in infants is inferred from eye gaze (Bailey & Plunkett, 2002, 2005; Swingley, 2009; Swingley & Aslin, 2000, 2002).

Although there are numerous studies that have used measures of PR accuracy with children, few have addressed the lexical restructuring hypothesis directly. Studies of PR accuracy have instead tended to focus either on the dimensionality of PR quality (e.g., Anthony et al., 2010, 2011) or on which tasks best differentiate between clinical and typically developing populations (e.g., Alt & Suddarth, 2012; Claessen & Leitao, 2012; Sutherland & Gillon, 2005). Within this study, measures of PR accuracy are used to test predictions that follow from the theoretical accounts discussed above.

#### *Measures of PR segmentation*

The second category of measures aims to assess the level of segmentation of children's representations. One of the most direct measures of PR segmentation is the gating task developed by Grosjean (1980). The rationale here is that if



children's PRs are less segmented than adults', they will need to hear more of the word before they are able to recognize it. This has been shown to be the case (Elliot, Hammer, & Evan, 1987; Walley, 1988; Walley, Michela, & Wood, 1995), providing support for the LRM, although more recent studies have questioned the reliability of the gating task for measuring PRs in young children (Sutherland & Gillon, 2005; Wesseling & Reitsma, 2001). Wesseling and Reitsma (2001) found gating to be a poor predictor of phonological awareness and to give inconsistent results at retest, while in Sutherland and Gillon's (2005) study, gating performance did not differentiate between children with specific language impairment and the control group.

Another way to look for evidence of PR segmentation is to investigate position effects in speech processing tasks. For example, while adults have been found to detect initial mispronunciations more easily than mispronunciations elsewhere in a word, children are less sensitive to the position of the mispronunciation (Cole & Perfetti, 1980; Walley, 1987). This finding has been interpreted as evidence that children's PRs are less segmented than adults', with children needing to focus on the whole word rather than being able to rely on the onset as adults appear to do (Metsala & Walley, 1998).

The tasks outlined above all seek to measure the segmentation of PRs in the most intuitive sense; that is, they assess the extent to which the representations are stored in terms of separable chunks (whether words are stored as a holistic whole word template, e.g., "/kæt/"; as an onset and a rime segment "/k/-/æt/"; or as three separate phonemes, "/k/-/æ/-/t/"). However, other authors have operationalized PR segmentation differently, in terms of the degree of specification of each segment. For example, in Storkel's (2002) strong restructuring account, a consonant-vowel-consonant (CVC) word from a sparse neighborhood (i.e., one with few phonological neighbors) like "boom" will be stored in terms of three segments from infancy, but the final segment /m/ will initially be coded coarsely as simply /nasal/. At this stage of development, the final segment of the word is indistinct from other nasals. It is only later in childhood when the final segment will be stored phonemically. Evidence in support of this strong restructuring account comes from similarity judgment tasks where children are asked to make classifications about interword similarity (Carroll & Myers, 2011; Storkel, 2002). The rationale behind similarity judgment tasks is that by finding out which types of words children classify as being similar, we can make inferences about the way that the words are stored within their lexicon.

Nonword repetition has also been used as a measure of PR segmentation (Coady & Aslin, 2004), although the extent to which a nonword repetition task taps into stored phonological representations is unclear. Although nonword repetition was traditionally thought of as being a pure measure of phonological short-term memory (PSTM) that is free from lexical influences, it has been argued that nonword repetition taps into both lexical and sublexical processes and can be used as a probe into children's stored representations (Coady & Aslin, 2004; Coady & Evans, 2008).

We have seen above that theoretical accounts of PR development make different predictions about if and how the accuracy and level of segmentation of children's PRs change over development. Previous studies have tended to focus on either

PR accuracy or PR segmentation and have not measured both within the same sample. This study has measured both aspects of PR quality concurrently alongside measures of both expressive and receptive vocabulary size.

Four of the PR tasks are receptive, eliminating the potential confound of speech output difficulties. The receptive PR tasks have also been designed to be as implicit as possible so that they do not require any explicit knowledge of phonemes. In this way, we were able to test some of the predictions made by the LRM and to investigate the relationships between the different PR measures.

## AIMS

The current study used measures of PR accuracy and segmentation in nursery children (aged 3 years, 6 months [3;6] to 4;6) to investigate the following questions:

1. Is there evidence in support of vocabulary-driven restructuring (in the form of a relationship between vocabulary size and PR segmentation)?
2. What is the level of segmentation of nursery children's PRs?
3. What are the core underlying capacities that determine performance on PR measures?

## METHOD

### *Participants*

Thirty-eight children were recruited from the nursery classes of three mainstream, state-run schools with low, medium, and high sociodemographic status (as indexed by the proportion of free school meals). Children's ages ranged between 3;6 and 4;6 with an average age of 4;1. There was an equal number of males and females. Children with known hearing difficulties (as reported by the teacher) and children on the special educational needs register for speech, language, or developmental issues were excluded from the study. Only children speaking English as their first language with at least one English-speaking parent/carer at home were included in the study. All children within the study spoke British English, although there were a variety of accents within the sample.

### *Procedure*

Children were tested individually in a quiet area within school over three sessions each lasting between 30 and 45 min. All tasks requiring a verbal response were audiorecorded for checking purposes.

The two repetition tasks, which were relatively short, were always administered in full, regardless of performance. For all other tasks, a stopping criterion was applied to avoid unnecessary testing: if after eight items the child was scoring at or below chance, the test was aborted at this point. This criterion affected five children on the initial sound task and one child on the mispronunciation reconstruction and detection tasks. Across all tasks corrective feedback was given for training items only.

Table 1. *Summary of measures of phonological representation (PR)*

Task	Description	Test Items	Type of PR Measure	Modality
Initial sound	Guess what the puppet wants from the initial phoneme?	24	Segmentation	Receptive
Mispronunciation reconstruction	Guess which word the puppet was trying to say?	24	Segmentation	Receptive
Mispronunciation detection	Did the puppet say the word correctly?	24	Accuracy	Receptive
Mispronunciation decision	Which of the two puppets said the word correctly?	24	Accuracy	Receptive
Word repetition	Repeat the words after the monster.	20	Accuracy	Expressive
Nonword repetition	Repeat the nonsense words after the monster.	20	Segmentation	Expressive

*Materials*

All experimental tasks used target words and distracters selected to be familiar to nursery children. For all tasks involving visual stimuli, the children were asked to name the pictures to check that they had identified them correctly; when children were unable to name the picture, the experimenter told them the correct name.

For each task, test trials were preceded by one or more practice items. All experimental tasks used two different orders to control for order effects, and the physical position of the target item in tasks using visual stimuli was randomized. The stimuli list for each task and detailed information on the protocol and selection of task materials are provided in the online-only supplementary materials.

*Background language measures*

Measures of the children’s expressive and receptive vocabulary were taken using the Renfrew Word Finding Vocabulary Test (Renfrew, 1995) and the British Picture Vocabulary Scale (BPVS; Dunn et al., 2009), respectively. The delivery of the BPVS test was slightly different from the standard method in that all children started from Set 1, regardless of age. This was to allow raw scores to be compared across 3- and 4-year-old children.

*Measures of phonological representation*

A battery of six measures was developed, which was designed to probe the level of segmentation and the accuracy of children’s phonological representations (Table 1). The battery includes four receptive measures (which do not require a verbal response) and two expressive measures.

### *Receptive PR measures.*

**INITIAL SOUND TASK.** In this novel task, children were asked to select which picture (out of a choice of four) they thought a puppet wanted after hearing it say an initial sound; for example, the child heard /p/ and chose from the target picture, paint, and three distracters: bed, toilet, and rug (see Figure 1a). The task was designed to measure PR segmentation with the rationale that children would only be able to identify the word from its initial sound if their PRs were segmented at least at the onset-rime level. The stimuli for the initial sound task were spoken live to the child by the experimenter (pretending to be the puppet).

This task is similar to the “phoneme identification” task that often appears in batteries testing explicit phonological awareness (e.g., Elbro & Petersen, 2004), where the child is shown pictures of objects and is asked, “Which one starts with . . . .” Both tasks require children to access their PRs given that they are “silent” tasks (where the tester does not provide a spoken example of the test words). The key difference between the two tasks is that in the initial sound task the child is not asked explicitly to operate on phonemes; rather they are simply asked to guess what the puppet wants. This distinction is important because it follows that while the phoneme identification task used in previous studies taps explicit PA, this novel task has the potential to tap the segmentation of PRs before PA emerges.

**MISPRONUNCIATION RECONSTRUCTION AND DETECTION TASK.** The two measures of mispronunciation detection and mispronunciation reconstruction were combined within one task for ease of administration. The detection component probes PR accuracy, but the reconstruction element probes the level of segmentation of children’s PRs.

Children were asked to identify whether or not a word was pronounced correctly (mispronunciation detection) by an on-screen alien. The stimuli were prerecordings of the experimenter saying the correct and incorrect forms of the word. Mispronunciations deviated from the target word by a single phoneme substitution or deletion. The rationale here is that children will only be able to detect minimally contrasting mispronunciations if they are storing the phonological word forms accurately. In contrast, if a child has “fuzzy” underspecified representations of words, we would expect them to score poorly on mispronunciation detection. Similar measures of mispronunciation detection used in previous studies have been shown to be reliable, valid measures of PR accuracy with the potential to identify children at risk of later literacy problems (Claessen et al., 2009).

For those words that were mispronounced by the puppet, the child was then asked to identify the target word (from a choice of four pictures) from the mispronounced form (mispronunciation reconstruction). For example, the child hears /pag/ and decides whether the puppet was trying to say pig (target), pen, dog, or bowl (see Figure 1b). While the target word differed from the mispronunciation by only a single phoneme substitution/addition, the distracters all involved at least two phoneme substitutions. It is assumed that success on this task requires the child to compare each of the four words with the mispronunciation and to select the one that is most phonologically similar. It is further assumed that the child will only be able to do this if his or her PRs of the words are segmented.

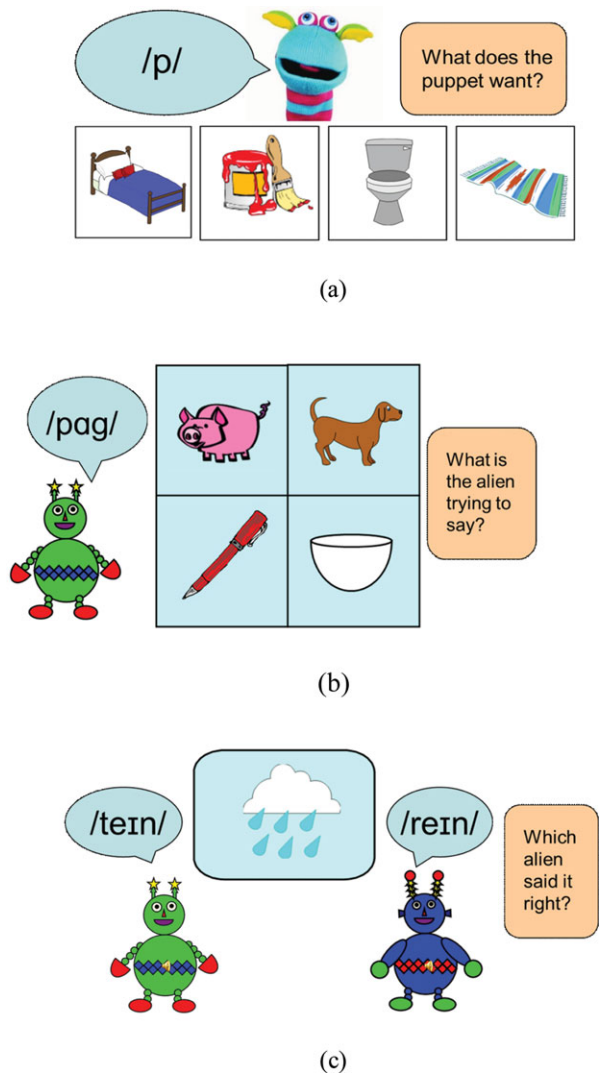


Figure 1. (Color online) Example items from three measures of phonological representation. In each case, the speech bubbles show what the child would hear (children were not shown the orthographic form of the stimulus). (a) Initial sound task: can the child implicitly link the initial segment (/p/) to the whole word (paint)? (b) Mispronunciation reconstruction task: does the child choose the word (pig) that shares the most phonemes with the mispronunciation (parg)? (c) Mispronunciation decision task: can the child identify which alien pronounces the word correctly?

The mispronunciation reconstruction component is a novel measure of PR segmentation that was developed as a way to probe the level of segmentation of children's PRs without requiring any explicit awareness of sound structure (PA). The mispronunciation reconstruction task is superficially similar to Fowler and Swainson's (2004) "correction of naming errors" task (a variant of articulation correction), where children are also asked to listen to a mispronunciation and identify which picture the character is trying to say. However, the two tasks are designed to measure different aspects of PR ability and, accordingly, involve different task demands. The correction of naming errors task aims to measure PR precision rather than segmentation and so does not manipulate phonological similarity between the stimuli and the distracters as the reconstruction task does. The emphasis in the correction of naming errors task is on how well the child corrects the word not on the child's ability to identify which word it is. The distracters are therefore more distant from the target word (varying in holistic features such as word length and prosody) than they are in the reconstruction task (where the distracters vary in terms of phonology alone).

**MISPRONUNCIATION DECISION TASK.** Children were asked to choose which of two pronunciations was the correct production for a pictured item (see Figure 1c). For example, the child hears two on-screen aliens try to say rain (/rein/). While one alien pronounces the word correctly, the other says /teim/. Again the stimuli were spoken and prerecorded by the experimenter. This task was included to complement the more commonly used mispronunciation detection task, because of the potential increase in sensitivity afforded by two alternative-choice methods over yes-no paradigms (MacMillan & Creelman, 2005).

Each of the four receptive PR tasks consisted of 24 items (see online-only supplementary materials for stimuli lists).

#### *Expressive PR tasks.*

**WORD AND NONWORD REPETITION TASKS.** In the word repetition task, children were asked to repeat 40 prerecorded CVC words (spoken by the experimenter). In the nonword repetition task, children were asked to repeat 40 CVC nonwords (again spoken by the experimenter and prerecorded). The nonword stimuli list was generated by changing the vowels in the word repetition stimuli list. The proportion of short versus long vowels or diphthongs was matched for the word and nonword stimuli. Children were only scored as correct if they pronounced the whole of the word/nonword correctly. For both tasks, the children were asked to repeat three training words (also CVC) before the testing phase began, to ensure that they had understood the task.

Word and nonword repetitions were included to allow us to investigate children's performance on expressive as well as receptive measures and to allow us to replicate Vance et al.'s (2005) finding that young children process words and nonwords in the same way. Word repetition has been used previously as a measure of PR accuracy, the idea being that in repetition, the child's existing representation is activated that in turn drives production of the corresponding motor program (e.g., Vance et al., 2005). It is assumed that for a child to correctly repeat a word,

the child's PR for that word must be stored accurately. Nonword repetition, in contrast, is traditionally considered to be a measure of PR segmentation. It is assumed that when children are asked to repeat a nonword stimulus for which they do not have an existing lexical representation, they need to mentally segment the word form into phonological chunks (e.g., phonemes or onset-rime segments) so that they can access the required motor programs and then reassemble them ready for production (Vance et al., 2005).

However, note that although word repetition is traditionally assumed to measure "top down" lexical level processes and nonword repetition is thought to involve "bottom up" segmental processes, Vance et al.'s (2005) findings suggest that young children may use segmental processes for repetition of words and nonwords. Coady and Evans (2008) also suggested that nonword repetition is not a purely segmental process; rather, nonword repetition performance is also affected by children's existing representations for known words. By including both word and nonword repetition within the battery, we are able to compare children's performance on the two measures and investigate whether or not the two measures seem to be measuring different processes. Although the word and nonword repetition tasks are given their traditional labels as measures of PR accuracy and segmentation within Table 1, we acknowledge there is ambiguity surrounding the degree to which lexical/sublexical processing comes into play within the two tasks. By including both tasks within the battery, using equivalent stimuli, we are able to compare children's performance on the two measures and investigate whether the two measures seem to be measuring different processes.

## RESULTS

### *Data preparation*

The raw scores for all experimental tasks were converted into the percentage of correct responses. The BPVS and word finding scores were kept in raw form because the number of items administered to each child is dependent on performance. For the mispronunciation reconstruction task, the percentage of correct responses was calculated for the mispronounced items only.

Five task scores were identified from the boxplot (Figure 2) as potential outliers, with values more than 1.5 times the interquartile range below the first quartile. Following inspection of the raw data, two children were excluded listwise and one participant's score on the nonword repetition task was substituted with the mean. There were five cases where children withdrew from a task having completed fewer than eight items. These missing scores were also substituted with the mean value for that task, as recommended by Field (2009, p. 654).

### *Description of performance on the PR tasks*

The descriptive statistics for performance on the six PR tasks and the two vocabulary measures are presented in Table 2. For the mispronunciation detection and decision tasks, both accuracy scores (% correct) and measures of sensitivity are reported. Values of  $d'$  were calculated for the mispronunciation detection



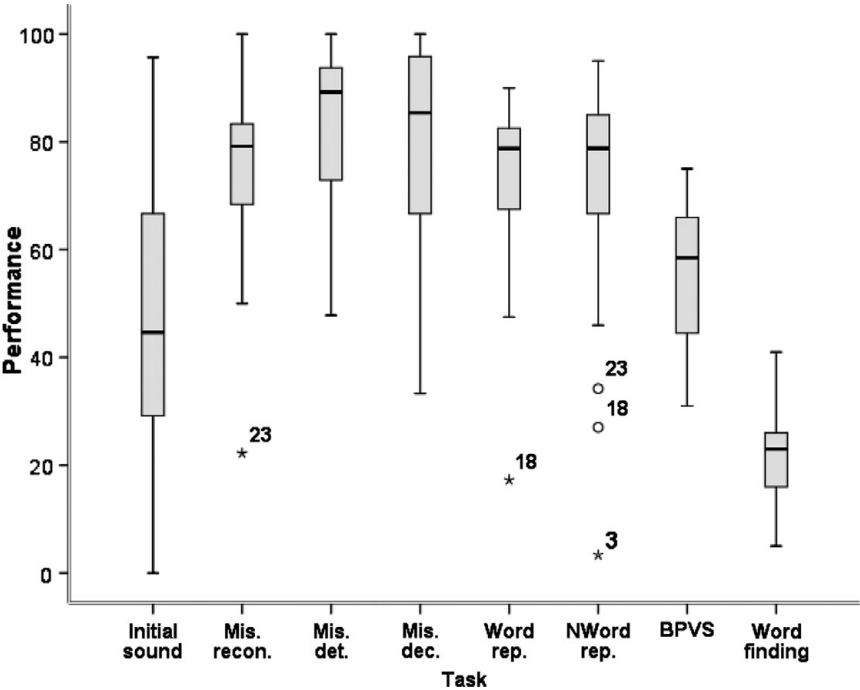


Figure 2. Boxplot of performance on the phonological representation tasks and vocabulary measures. All scores are plotted as the percentage of items correct except for the British Picture Vocabulary Scale and the word finding scales (presented as raw scores). Moderate ( $1.5 \times \text{IQR} \pm \text{third/first quartile}$ ) and extreme outliers ( $3 \times \text{IQR} \pm \text{third/first quartile}$ ) are denoted by circles and stars, respectively, and are labeled by participant number.

and decision tasks to separate the child’s sensitivity to mispronunciations from any response bias that the child might have, in line with signal detection theory (MacMillan & Creelman, 2005).

*Summary data for the receptive PR tasks*

The four receptive PR tasks each consisted of 24 items. According to the binomial distribution, a score of 16 or above on the detection and decision tasks (which both involve two choices) is considered to be significantly above chance. For the initial sound and mispronunciation reconstruction tasks (which both involve four choices), a score of 10 or above is considered to be significantly above chance. The percentages of children scoring above chance (shown in Table 2) indicate that the majority of these preschool children were capable of detecting and making judgments about mispronunciations within one-syllable words, reconstructing a target word from its mispronunciation and identifying a whole word target from its initial sound (although they were not at ceiling level on any task).

Table 2. *Performance on phonological representation and expressive and receptive vocabulary measures (after exclusion of outliers and missing cases)*

Task	N	Mean	SD	% Children Above Chance
Initial sound (% correct), chance level = 25%	36	49.80	23.65	68.42
Mispronunciation (% correct)				
Reconstruction				
Chance level = 50%	36	78.15	10.11	100
Detection				
Chance level = 50%	36	84.32	13.76	92.10
<i>d'</i> , sensitivity, chance = 0		2.18	0.94	
Decision				
Chance level = 50%	33	84.66	15.71	85.71
<i>d'</i> , sensitivity, chance = 0		1.62	0.76	
Word repetition (% correct)	36	76.72	9.44	
Nonword repetition (% correct)	35	76.64	11.85	
BPVS (raw score)	34	56.50	12.99	
Word finding (raw score)	36	22.33	6.90	

Note: BPVS, British Picture Vocabulary Scale.

*Effect of mispronunciation type*

Metsala and Walley (1998) proposed that differential mispronunciation detection performance for items involving initial substitutions versus mispronunciations elsewhere in the word is indicative of the level of PR segmentation. If children’s PRs are restructured by vocabulary growth, we might expect children with higher vocabularies to detect mispronunciations in initial position more readily and for children with lower vocabularies to be unaffected by the position of the mispronunciation. We might also expect to see an interaction between vocabulary size and position effects for the mispronunciation decision and mispronunciation reconstruction tasks. To test these predictions, the difference between the percentage of items correct where the mispronunciation was in the initial position minus the percentage of items correct where the mispronunciation was in the medial or final position was calculated for each child (performance on items involving deletion of consonants from consonant clusters was omitted from this analysis) for each of the three mispronunciation tasks. Children were then separated into two groups of low and high expressive vocabulary (word finding score) using a median split. On the mispronunciation detection task, the low vocabulary group performed equally well on items involving initial substitutions and substitutions in other positions (i.e., medial and final),  $t(20) = 0.27, p = .40$ , while the high vocabulary group performed better on initial substitutions than on medial/final substitutions,  $t(14) = 2.03, p = .031$ . On the mispronunciation reconstruction task, both high vocabulary and low vocabulary groups performed equally well on initial mispronunciation items and items with the mispronunciation in the medial/final position, low:  $t(20) = 1.33, p = .099$ ; high:  $t(14) = 1.41, p = .09$ . On the mispronunciation

decision task, there was no effect of mispronunciation position for either the low or high vocabulary group, low:  $t(17) = 0.86, p = .40$ ; high:  $t(17) = 0.00, p = 1.00$ .

When we group the children according to receptive vocabulary (again using a median split but with the BPVS scores), we find that on the mispronunciation detection task children, there are no position effects for either the low or high vocabulary groups, low:  $t(16) = 0.54, p = .30$ ; high:  $t(18) = 1.45, p = .08$ . On the mispronunciation reconstruction task, there is a significant effect of position for the high vocabulary group,  $t(18) = 2.05, p = .028$ , but not for the low vocabulary group,  $t(16) = 0.62, p = .54$ . Once again, no position effects were found for the mispronunciation decision task, low:  $t(13) = 1.54, p = .15$ ; high:  $t(18) = -0.66, p = .52$ . The lack of a position effect on the mispronunciation decision task could be because many children scored highly on this task, with the children with higher vocabularies performing especially well. Because of this, the difference between performance on mispronunciations in initial versus medial or final position tended to be small, especially for the children with higher vocabularies.

### *Comparison of word and nonword repetition performance*

Word and nonword repetition tasks were compared, in an attempt to replicate Vance et al.'s (2005) finding that 3-year-olds perform equally well on the two types of repetition tasks, while 4-year-old children find it easier to repeat words than nonwords. Analysis of variance was carried out on the word and nonword repetition percentage scores with children grouped according to age (3 years,  $N = 14$ ; 4 years,  $N = 24$ ). No significant main effect of task was found,  $F(1, 33) = 0.003, p = .96$ , indicating that children performed equally well on the word and nonword repetition tasks. A marginally significant main effect of age was found,  $F(1, 33) = 4.07, p = .054$ , with 4-year-olds performing better than 3-year-olds. There was no significant interaction of task with age,  $F(1, 33) = 0.003, p = .96$ .

### *Factor analysis*

Factor analysis was conducted on the PR task scores to investigate the underlying capacities involved in PR performance. The PR tasks were designed to measure the accuracy and level of segmentation of children's PRs. We would, therefore, expect at least one factor to emerge, representing PR quality. If the lexical restructuring hypothesis holds, the predicted factor(s) would also be expected to correlate with measures of receptive and expressive vocabulary (Metsala & Walley, 1998). In order to test this prediction, factor analysis was performed on six variables representing performance on the six PR task scores: initial sound, mispronunciation reconstruction, mispronunciation detection, mispronunciation decision, word repetition, and nonword repetition (with the sensitivity measure  $d'$  rather than percentage accuracy used for the mispronunciation detection and decision tasks to control for the effect of response bias). The measures of receptive and expressive vocabulary were not included in the factor analysis so that they could be used as predictors of the factor scores, thus providing an independent test of the interpretation of the factor model.

Table 3. *Correlation matrix for the phonological representation scores, age, and background language measures*

Measure	1	2	3	4	5	6	7	8	9
1. Initial sound	—								
2. Mispron. reconstruction	.42**	—							
3. Mispron. detection, <i>d'</i> (sensitivity score)	.45**	.45**	—						
4. Mispron. decision, <i>d'</i> (sensitivity score)	.25	.46**	.58**	—					
5. Word repetition	.20	.28	.061	-.21	—				
6. Nonword repetition	.26	.33*	.090	.17	.36*	—			
7. BPVS (raw score)	.29*	.15	.44**	.39*	.045	-.047	—		
8. Word finding	.51**	.35*	.49**	.35*	.14	.13	.61**	—	
9. Age	.53**	.34*	.27	.095	.11	.22	.17	.45**	—

Note: All scores are percentage correct unless otherwise stated. BPVS, British Picture Vocabulary Scale.

\* $p < .05$ . \*\* $p < .01$ .

The preliminary stage of the analysis involved examination of the correlations between task scores as reported in Table 3. The four receptive PR tasks designed to measure PR accuracy (the mispronunciation detection task and the mispronunciation decision task) and PR segmentation (the initial sound task and mispronunciation reconstruction task) each correlated significantly with at least one of the vocabulary measures, in line with the LRM. The expressive repetition tasks, however, did not correlate with either vocabulary measure.

Significant nonnormality was found for the sensitivity scores for the mispronunciation detection,  $D(36) = 0.16, p = .03$ , and mispronunciation decision tasks,  $D(33) = 0.16, p = .03$ , and for the word repetition accuracy scores,  $D(36) = 0.16, p = .02$ . This motivated use of the principal axis factoring method that is preferable for use in cases of significant nonnormality (Fabrigar, Wegener, MacCallum, & Strahan, 1999). The Kaiser–Meyer–Olkin measure (0.63) was above the acceptable threshold of 0.5 (Field, 2009, p. 659), confirming sampling adequacy for the analysis. The Bartlett test of sphericity was significant,  $\chi^2(15) = 46.98, p < .0005$ , indicating that the correlations between variables were sufficiently high for factor analysis to be conducted. Two factors were retained in line with Kaiser’s criterion of eigenvalues  $> 1$  (Kaiser, 1960). The scree plot (see online-only Supplementary Materials) was also inspected, confirming that two factors would yield the optimal solution. Together the factors explained 48.55% of the variance.

Direct oblimin rotation was employed initially to check the orthogonality of the factors. The correlation between the factors was found to be nonsignificant ( $r = .23, p = .18$ , two tailed) motivating employment of varimax orthogonal rotation (see Figure 3) for ease of interpretation. Table 4 shows the factor matrix after varimax rotation with loadings  $> 0.32$  shown in bold (using the rule of thumb that loadings over 0.32 have substantive importance; e.g., Tabachnick & Fidell, 2007).

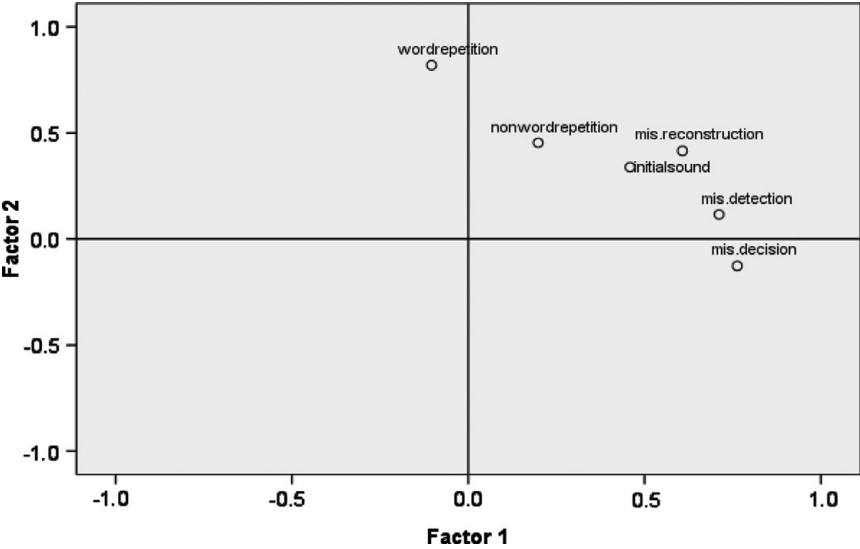


Figure 3. Factor plot in rotated factor space, where Factors 1 and 2 are proposed to represent phonological representation quality and phonological processing abilities, respectively.

Table 4. *Summary of exploratory factor analysis results for performance on the phonological representation tasks*

	Factor 1	Factor 2
Mispronunciation decision	<b>0.76</b>	−0.13
Mispronunciation detection	<b>0.71</b>	0.12
Mispronunciation reconstruction	<b>0.61</b>	<b>0.42</b>
Initial sound	<b>0.46</b>	<b>0.34</b>
Word repetition	−0.10	<b>0.82</b>
Nonword repetition	0.20	<b>0.45</b>
Eigenvalues	1.71	1.20
% of variance	28.61	19.93

*Note:* The factor matrix after varimax rotation with loadings >0.32 are in bold.

If low loadings are ignored (<0.32), the mispronunciation decision and detection tasks load onto Factor 1 only, the two repetition tasks load onto Factor 2 only, and the initial sound and mispronunciation reconstruction tasks load onto both factors. In an attempt to interpret the meaning of the two factors, the communalities and differences between the demands of the six tasks were considered. The two tasks loading highly onto Factor 1, mispronunciation detection and decision, both test children’s ability to identify whether a word is spoken correctly. The two tasks loading highly onto Factor 2 both test the child’s ability to listen to

a speech signal and repeat it, but the interpretation cannot simply relate to production because the initial sound and mispronunciation reconstruction tasks load moderately onto both factors. The two factors, therefore, need to be interpreted in terms of two skills/abilities that are both required for the initial sound and mispronunciation reconstruction tasks but that discriminate between the repetition tasks and the mispronunciation detection/decision tasks. While Factor 1 is proposed to represent PR quality, two alternative possibilities for the interpretation of Factor 2 are proposed: PSTM and segmental processing (see Discussion).

If the lexical restructuring hypothesis is correct, we would expect the factor scores for the PR quality factor to correlate significantly with vocabulary. This prediction is borne out in a correlational analysis with Factor 1 correlating moderately with both measures of vocabulary (word finding:  $r = .47$ ,  $p = .002$ ; BPVS:  $r = .42$ ,  $p = .005$ ). This relationship remains significant when controlling for age (word finding:  $r = .39$ ,  $p = .01$ ; BPVS:  $r = .39$ ,  $p = .01$ ). While the correlation between expressive vocabulary and Factor 2 was marginally significant ( $r = .23$ ,  $p = .09$ ), Factor 2 did not correlate significantly with receptive vocabulary ( $r = .041$ ,  $p = .41$ ). Factor 2 did not correlate significantly with either vocabulary measure when controlling for age (word finding:  $r = .13$ ,  $p = .22$ ; BPVS:  $r = -.002$ ,  $p = .50$ ). Unless stated otherwise, all the significance values above are one tailed because the corresponding a priori predictions were directional (i.e., all relationships were expected to be positive).

## DISCUSSION

In this study we have used four receptive tasks and two repetition tasks to investigate the accuracy and level of segmentation of nursery children's phonological representations. While a number of studies have explored the development of phonological representations, they have tended to look at either PR accuracy (e.g., Anthony et al., 2010; Claessen et al., 2009) or PR segmentation (e.g., Foy & Mann, 2009), but not both. Furthermore, most studies have been restricted to children who have already begun reading instruction, or have used measures that probe sensitivity to phones rather than phonemic representation. The current study has measured PR accuracy and segmentation concurrently within the same sample and has provided evidence of lexical restructuring in preliterate children as young as 3;6.

### *Evidence of PR segmentation in preliterate children*

Performance on the initial sound task provides evidence that the PRs of nursery children (aged 3;6 to 4;6) are segmented at least at the onset-rime level, with children performing well above floor levels on this task. If children's PRs were stored holistically, we would not expect them to be able to relate an isolated initial sound segment to their whole-word template of the target item. It is assumed that for success on this task, the initial sound needs to be separable from the rest of the word within the child's phonological representation. It is worth noting that the scores were not at ceiling for this task and that there was considerable variation in performance across participants. This indicates that children do not

find linking an initial sound to its whole-word referent a trivial task. Scores showed the highest correlation with age, suggesting that this skill is still developing within the preschool period.

All of the children performed above chance on the mispronunciation reconstruction task, which requires sensitivity to the number of shared segments between the stimulus and the target word. If we assume that this sensitivity requires (or is at least facilitated by) segmented PRs, then the high levels of performance observed within this sample provide further support for PR segmentation in nursery children. It is possible that children might be able to carry out the task to some extent using holistic representations, with children selecting the word that is most similar overall rather than in terms of shared segments. However, it would be difficult to explain why the mispronunciation reconstruction task clusters with the initial sound task if they are not both measures of segmentation.

### *Evidence in support of vocabulary driven restructuring*

We have presented evidence consistent with the view that children's phonological representations are segmented before they enter formal education. In order to distinguish between the accessibility and emergent viewpoints, we need to know if the level of segmentation of children's PRs is dependent on vocabulary size, while controlling for age. If children's PRs have been segmented from infancy, we would not expect measures of PR segmentation to correlate with vocabulary. If, in contrast, PRs only become segmented gradually as the child's lexicon expands, we would expect to see a significant relationship between vocabulary size and measures of PR segmentation. All four receptive PR measures correlated significantly with expressive vocabulary, with three out of the four correlating with receptive vocabulary. This is consistent with the LRM's proposition that children's representations of words will increase in terms of both accuracy and segmentedness under the pressure of a growing vocabulary (Metsala & Walley, 1998). Further evidence in support of the LRM is that position effects were found on the mispronunciation detection task for children with high but not low expressive vocabularies and on the mispronunciation reconstruction task for children with high but not low receptive vocabularies. This is in line with previous studies that show adults to be more sensitive to position effects in mispronunciation detection than are children (Cole & Perfetti, 1980; Walley, 1987).

The results are also consistent with PRIMIR, which predicts that phonemic representation emerges gradually as children learn more words and regularities appear across the word form plane (Werker & Curtin, 2005). Although PRIMIR does not predict PR accuracy to be vocabulary dependent (children's representations are assumed to be very detailed from the beginning), the relationship between vocabulary and measures of PR accuracy could be explained in terms of vocabulary mediating the lensing effect of the child's level of development. Within PRIMIR, the child's developmental level (something that may potentially be indexed by vocabulary size) acts as a filter that determines the degree to which certain levels of information are processed during certain tasks. In summary, the current findings support the idea of emerging phonemic representation in preliterate children



driven by vocabulary growth and are consistent with both the LRM (Metsala & Walley, 1998) and PRIMIR (Werker & Curtin, 2005).

### *What is the relationship between the different PR measures?*

The six PR tasks were devised to provide measures of the accuracy and level of segmentation of preschool children's PRs. It was predicted that factor analysis would yield at least one factor representing PR quality that would correlate significantly with vocabulary in line with the LRM. This prediction was confirmed, with two factors emerging: one representing PR quality and a second factor that lends itself to one of two interpretations: PSTM or segmental processing. Interpretation of the two factors was made on the basis of the demands associated with each task.

When labeling the factors, we considered which task demands were shared by the four receptive tasks but not by the two repetition tasks. The key property that the four receptive tasks share is that they all require the child to compare the heard stimulus with the stored representations for the words shown in the picture choices. It follows that they all require accurate PRs, yielding the PR quality label for Factor 1. The key property that the two segmentation tasks share with the nonword repetition is that they all require the child to hold the stimulus in PSTM while some degree of segmental processing takes place. For the initial sound task, children need to retain the heard segment in PSTM (s) while they access the PRs for the four picture choices. They then need to compare the segment with each of the initial segments for the four word choices in search of a match. For the mispronunciation reconstruction task, children need to hold the mispronunciation in PSTM, and access their existing PRs for each picture and then for the closest segmental match. For the nonword repetition task, children need to retain the stimulus in memory, break it down into its constituent segments, and then reassemble it into the required motor plans (e.g., Vance et al., 2005). The mispronunciation detection and decision tasks are unlikely to require segmental processing, with children making judgments based on a simple search of their lexicon to see if the stimulus matches an existing entry. It is assumed here that the mispronunciation detection and decision tasks can be completed by holistic rather than segmental comparisons (Metsala & Walley, 1998) and are, therefore, less taxing in terms of phonological processing.

Given that nonword repetition and the two receptive PR segmentation tasks are likely to place the greatest demands on PSTM, it therefore makes sense to label the factor onto which these three tasks load as PSTM. However, we need to consider how word repetition fits in with this interpretation and explain why a task that is generally considered to be predominantly lexical (rather than sublexical) might load onto PSTM but not PR quality. That word repetition clusters with nonword repetition and does not load onto PR quality supports the suggestion made by Vance et al. (2005) that younger children adopt a sublexical bottom-up strategy to repetition. Although Vance et al. (2005) found evidence of lexical influences on repetition in 4-year-olds, the children in our sample were 4;6 and younger, and it may be that top-down effects on word repetition only begin to appear between 4;6 and 5;0. If this is the case, then both word and nonword repetition tasks in young children may be measuring sublexical rather than lexical processing.

It is also worth noting that the correlation between Factor 2 and vocabulary was not as high as we might expect given the well-established link between PSTM and vocabulary that can be found within the literature (e.g., Gathercole, Service, Hitch, Adams, & Martin, 1999; Gathercole, Willis, Emslie, & Baddeley, 1992). While the two receptive segmentation tasks both correlated significantly with expressive vocabulary, the two repetition tasks did not. This discrepancy is likely because the CVC words used were not sufficiently complex to be sensitive to this association. Gathercole and Adams (1993) found a significant correlation between repetition and vocabulary for words of two syllables and longer, and nonwords of three syllables and longer, but not for shorter items. It is also worth noting that the children taking part in this study were younger than the children within many of the studies where an association between PSTM and vocabulary has been shown (e.g., Gathercole et al., 1992, 1999). It is possible that the association observed between PSTM and vocabulary size within older samples may be driven at least to some extent by literacy.

An alternative interpretation of Factor 2 is that the factor does not measure PSTM but does measure some other element of segmental processing. We suggest that within the current study, the repetition stimuli were not complex enough to make significant demands on PSTM but were complex enough to engage other aspects of segmental processing. One possibility is that the repetition tasks were sensitive to the coarseness of the final segment within the word. According to Storkel's (2002) strong restructuring account, children represent the ends of words from sparse neighborhoods in less detail than the preceding segments; for example, "boom" might be stored as /b-u, -nasal/. In order to verify this possibility, further work is needed to explore the effects of neighborhood density on repetition performance and to investigate whether the factor structure changes when repetition measures involving more complex stimuli (i.e., multisyllabic words/nonwords) are used.

### *Conclusions*

This study has used novel implicit measures of PR segmentation alongside measures of PR accuracy and vocabulary size to provide clear evidence of restructuring in children aged 3;6 to 4;6. The findings have two key theoretical implications. First, the findings suggest that children's PRs are emergent, with PRs becoming increasingly accurate and segmented over development. This is in contrast with the accessibility viewpoint that proposes that PRs are phonemic from infancy. Second, the findings suggest that PR restructuring is driven predominantly by vocabulary growth, as predicted by the LRM (Metsala & Walley, 1998) with the PR quality factor correlating significantly with vocabulary even when controlling for age. Although it is possible that this restructuring may be driven at least in part by literacy experience (c.f. Ziegler & Goswami, 2005), that restructuring was found within this sample of children who have not yet received formal reading instruction suggests that it is vocabulary growth that plays the key role in the restructuring of children's phonological representations. The results are consistent with both the LRM (Metsala & Walley, 1998) and the PRMIR framework (Werker & Curtin, 2005).

## ACKNOWLEDGMENTS

We thank the children, parents, and teachers who were involved in this study. This work was supported by the Economic and Social Research Council (ES/J500094/1).

## SUPPLEMENTARY MATERIAL

To view the supplementary material for this article, please visit <http://dx.doi.org/10.1017/S0142716415000338>.

## REFERENCES

- Alt, M., & Suddarth, R. (2012). Learning novel words: Detail and vulnerability of initial representations for children with specific language impairment and typically developing peers. *Journal of Communication Disorders*, 45, 84–97.
- Anthony, J. L., Aghara, R. G., Solari, E. J., Dunkelberger, M. J., Williams, J. M., & Liang, L. (2011). Quantifying phonological representation abilities in Spanish-speaking preschool children. *Applied Psycholinguistics*, 32, 19–49.
- Anthony, J. L., Williams, J. M., Aghara, R. G., Dunkelberger, M., Novak, B., & Mukherjee, A. D. (2010). Assessment of individual differences in phonological representation. *Reading and Writing*, 23, 969–994.
- Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, 17, 1265–1282.
- Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, 32, 159–173.
- Carroll, J. M., & Myers, J. M. (2011). Spoken word classification in children and adults. *Journal of Speech, Language and Hearing Research*, 54, 127–147.
- Casslerly, E., & Pisoni, D. B. (2010). Speech perception & production. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 629–647.
- Claessen, M., Heath, S., Fletcher, J., Hogben, J., & Leitão, S. (2009). Quality of phonological representations: A window into the lexicon? *International Journal of Language and Communication Disorders*, 44, 121–144.
- Claessen, M., & Leitao, S. (2012). The relationship between stored phonological representations and speech output. *International Journal of Speech–Language Pathology*, 14, 226–234.
- Coady, J. A., & Aslin, R. N. (2004). Young children’s sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*, 89, 183–213.
- Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communication Disorders*, 43, 1–40.
- Cole, R. A., & Perfetti, C. A. (1980). Listening for mispronunciations in a children’s story: The use of context by children and adults. *Journal of Verbal Learning and Verbal Behavior*, 19, 297–315.
- Diehl, R. H., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149–179.
- Dunn, L. M., Dunn, D. M., Sewell, J., Styles, B., Bryzaska, B., Shamsan, Y., et al. (2009). *The British Picture Vocabulary Scale*. London: GL Assessment.
- Elbro, C., & Petersen, D. K. (2004). Long-term effects of phoneme awareness and letter sound training: An intervention study with children at risk for dyslexia. *Journal of Educational Psychology*, 96, 660–670.

- Elliot, L. L., Hammer, M. A., & Evan, K. E. (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. *Perception & Psychophysics*, 42, 150–157.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272–299.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15, 399–402.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.
- Fowler, A. E. (1991). How early phonological development might set the stage for phonological awareness. In S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 97–117). Hillsdale, NJ: Erlbaum.
- Fowler, A. E., & Swainson, B. (2004). Relationships of naming skills to reading, memory, and receptive vocabulary: Evidence of imprecise phonological representations of words by poor readers. *Annals of Dyslexia*, 54, 247–280.
- Foy, J. G., & Mann, V. A. (2009). Effects of onset density in preschool children: Implications for development of phonological awareness and phonological representation. *Applied Psycholinguistics*, 30, 339–361.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, 13, 361–377.
- Gathercole, S. E., & Adams, A. M. (1993). Phonological working memory in very young children. *Developmental Psychology*, 29, 770–778.
- Gathercole, S. E., Service, E., Hitch, C. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology* 13, 65–77.
- Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28, 887–898.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, 267–283.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *National Review of Neuroscience*, 8, 393–402.
- Juszyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3–28.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. S., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1–33). Ann Arbor, MI: University of Michigan Press.
- Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua*, 120, 1898–1909.
- MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.
- Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- McNeill, B. C., & Hesketh, A. (2010). Developmental complexity of the stimuli included in mispronunciation detection tasks. *International Journal of Language and Communication Disorders*, 45, 72–82.
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: Erlbaum.
- Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, 24, 45–64.
- Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, 32, 120–132.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustic Society of America*, 119, 2382–2393.
- Renfrew, C. (1995). *Word Finding Vocabulary Test*. London: Speechmark.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131–141.
- Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., et al. (1996). Localization of grasp representations in humans by PET: 1. Observation versus execution. *Experimental Brain Research*, 111, 246–255.
- Storkel, H. L. (2002). Restructuring similarity neighborhoods in the developing mental lexicon. *Journal of Child Language*, 29, 251–274.
- Sutherland, D., & Gillon, G. T. (2005). Assessment of phonological representations in children with speech impairment. *Language, Speech, and Hearing Services in Schools*, 36, 294–307.
- Swan, D., & Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology*, 66, 18–41.
- Swingle, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, 60, 252–269.
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147–166.
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13, 480–484.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). London: Pearson.
- Treiman, R., & Breaux, A. (1982). Common phoneme and overall stimulating relations among spoken syllables: Their use by children and adults. *Journal of Psycholinguistic Research*, 11, 569–597.
- Vance, M., Stackhouse, J., & Wells, B. (2005). Speech-production skills in children aged 3–7 years. *International Journal of Language and Communication Disorders*, 40, 29–48.
- Walley, A. C. (1987). Young children's detections of word-initial and -final mispronunciations in constrained and unconstrained contexts. *Cognitive Development*, 2, 145–167.
- Walley, A. C. (1988). Spoken word recognition by young children and adults. *Cognitive Development*, 3, 137–165.
- Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, 13, 286–350.
- Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, 16, 5–20.
- Walley, A. C., Michela, V. L., & Wood, D. R. (1995). The gating paradigm: Effects of presentation format on spoken word recognition by children and adults. *Perception & Psychophysics*, 57, 343–351.

- Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental model of speech processing. *Language Learning and Development, 1*, 197–234.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary. *Infancy, 3*, 1–30.
- Wesseling, R., & Reitsma, P. (2001). Preschool phonological representations and development of reading skills. *Annals of Dyslexia, 51*, 203–229.
- Ziegler, J. C., & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29.